Simulation Exercise: Investigation of exponential distribution in R and comparison of it with Central Limit Theorem

Jae Kwon

March 3, 2017

Overview

In this project the exponential distribution will be investigated in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. The distribution of averages of 40 exponentials is investigated, and thousand simulations were done.

Simulation

Variables:

```
lambda <- 0.2
exp <- 40

mu <- 1/lambda
sig <- 1/lambda
```

Simulation:
Run simulation 1000 times with size of 40 from an Exp($\mu = \frac{1}{0.2}$, $\sigma = \frac{1}{0.2}$) distribution. Mean is calculated for each simulation.

```
smean= NULL
for (i in 1 : 1000) smean = c(smean, mean(rexp(exp, lambda)))
```

Result

List of variables:

```
smu <- mean(smean)
diff_m <- abs(mu-smu)
svar <- var(smean)
tvar <- sig^2/exp
diff_v <- abs(svar-tvar)

library(ggplot2)
data <- data.frame(smean)
```

Simulation(sample) mean:

```
smu
```

```
## [1] 5.005954
```

Theoretical mean:

```
mu
```

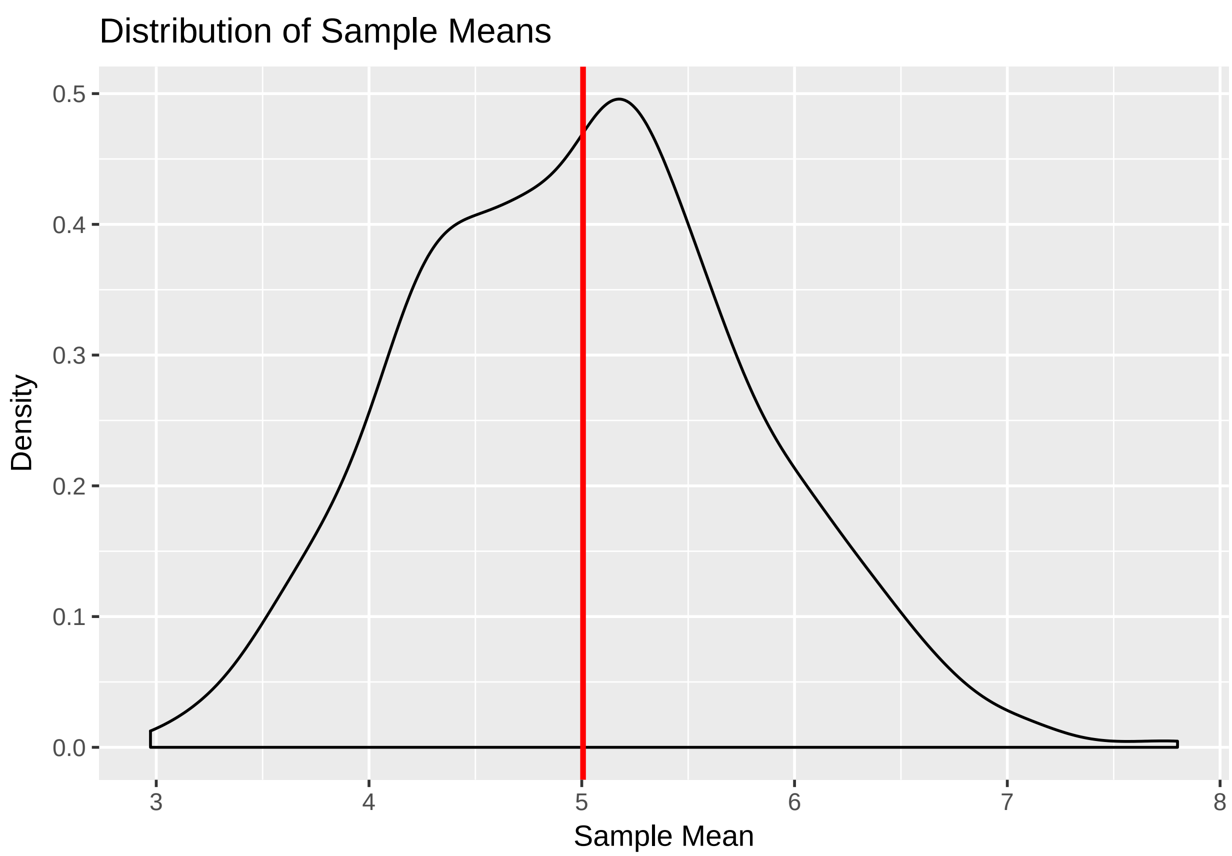
```
## [1] 5
```

Comparison:
Difference between simulation(sample) mean and theoretical mean:

```
diff_m
```

```
## [1] 0.005953662
```

```
ggplot(data, aes(smean)) + geom_density() +
       geom_vline(xintercept=mean(smean), size = 1, color = 'red') +
       xlab('Sample Mean') +
       ylab('Density') +
       ggtitle('Distribution of Sample Means')
```



Simulation(sample) variance:

```
svar
```

```
## [1] 0.6375963
```

Theoretical variance:

```
tvar
```

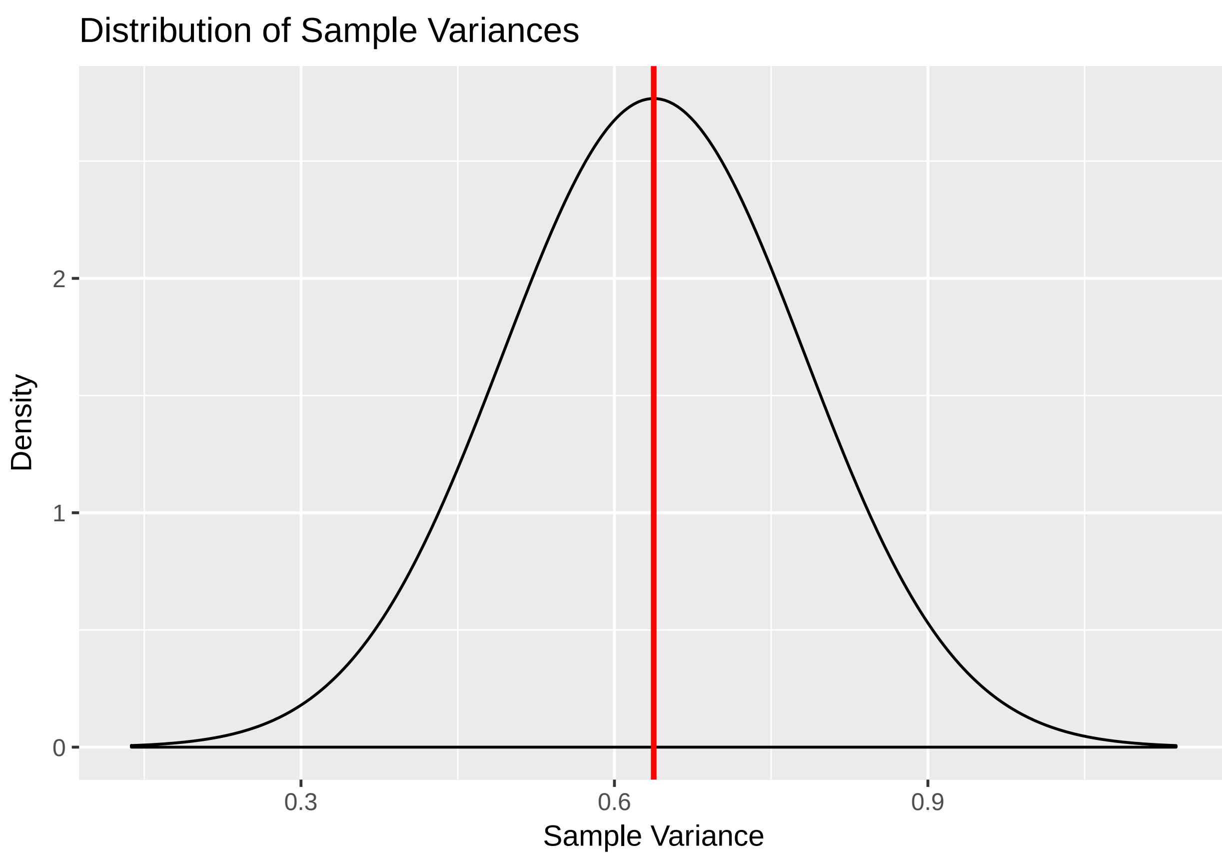
```
## [1] 0.625
```

Comparison:
Difference between simulation(sample) variance and theoretical variance:

```
diff_v
```

```
## [1] 0.01259626
```

```
ggplot(data, aes(svar)) + geom_density() +
       geom_vline(xintercept=mean(svar), size = 1, color = 'red') +
       xlab('Sample Variance') +
       ylab('Density') +
       ggtitle('Distribution of Sample Variances')
```

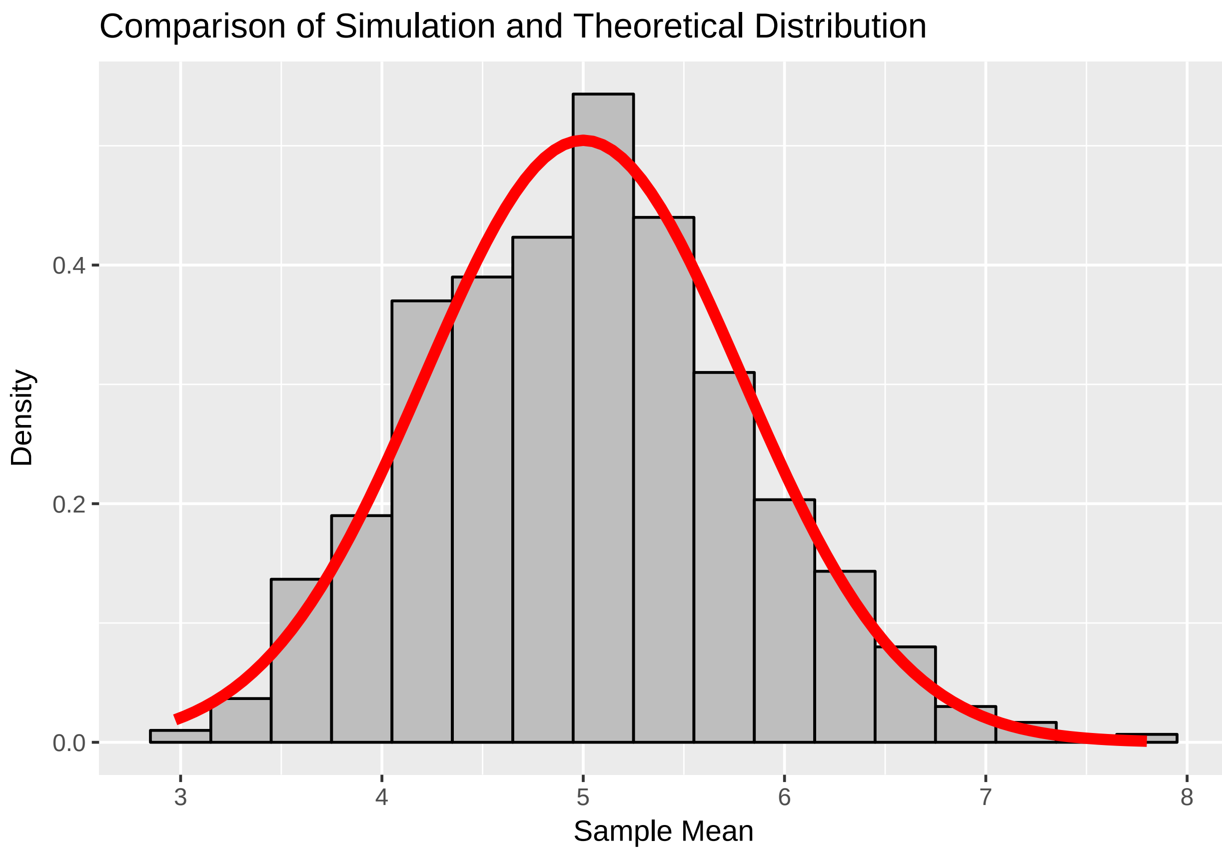


Distribution:

```
data <- data.frame(smean)

ggplot(data, aes(x = smean)) +
       geom_histogram(binwidth = 0.3, aes(y = ..density..), color = 'black', fill = 'gray') +
       stat_function(geom = "line",
                     fun = dnorm, args = list(mean = mu, sd = sqrt(tvar)),
                     color = 'red',
                     size = 2) +
       xlab('Sample Mean') +
       ylab('Density') +
```

```
    ggtitle('Comparison of Simulation and Theoretical Distribution')
```



The red line is the theoretical data, and the histogram is the distribution of the samples. The plot shows that the red line overlays on the histogram of the samples. Therefore, the sample distribution is roughly normal.